

The Value Of Information Sharing In Capacity Allocation In Serial Service Operations

Jin Kyung Kwak, Ewha School of Business, South Korea

ABSTRACT

In this study, we investigate the value of information sharing in serial service operations. When services are offered sequentially in two stages, we may use the demand information from the previous period in assigning servers at each stage. This study compares an information-based policy with a basic policy for capacity allocation in serial service operations in order to explore the value of information shared between the two stages. Among several possible candidates for an information-based policy, we chose an assigning rule such that the number of servers at the latter stage is determined by the number of customers served out of the prior stage in the previous period. On the other hand, the basic policy is designated as an assigning rule such that the two stages have the same constant number of servers through all periods. Assuming independent and identically distributed Normal demands with various parameters, we conducted computational experiments to compute the cost savings from using the information-based policy over the basic policy. The cost of using each policy includes the labor cost and the waiting cost. The results show that the cost savings of information sharing are relatively low and that the value of information sharing increases with demand variability or with unit waiting cost. These results give us some managerial insights on capacity allocation in serial service operations.

Keywords: Information Sharing; Capacity Allocation; Serial Service Operations

1. INTRODUCTION

There have been extensive studies on the value of information sharing in manufacturing supply chains. Particularly, in a serial supply chain, suppliers can reduce inventory costs by getting immediate end-customer demand information from retailers (Gavimani et al. 1999; Lee et al. 2000). In this study, by applying the similar research technique to that of studying supply chains, we pursue to investigate the value of information sharing in assigning the labor capacities for two-stage service operations. In other words, what if the two stages of service operations share the demand information together?

We may think of two-stage service operations in many businesses. For example, in hospitals, you may consult with a doctor first and then go to a pharmacist to get medications. Or you may see a nurse first and then get medical service from doctors afterwards. In financial loan services, there is an evaluation step and then consulting step. There may be many other examples in service settings, and we want to explore how information sharing between the two stages can help allocate the labor capacities.

There have been many studies on employee scheduling in service industries. Alfares (2004) classified the solution techniques for scheduling into ten categories and reviewed related literatures. Due to the complexity of this type of problems, many papers adapted simulation methods and proposed some heuristics (Zhu and Sherali 2009; Song and Huang 2008; and so on). Many existing scheduling literatures consider capacity allocation in services when facing uncertain demand, but any of them do not deal with serial service operations. Therefore, this study is the first to investigate the value of information sharing in serial service operations.

In this study, we modeled serial service operations as a two-stage supply chain so that we can analyze the value of information sharing between two stages. Demand uncertainty seems to highly influence on workforce planning

(Cappelli 2009), and so the process of forecasting the pattern of workload and managing the workforce to the staffing needs by time period can lead to very large cost reductions compared to stable staffing approaches where the number of employees held constant throughout the full period (Browne 1997). In this paper, we observe how information sharing leads to cost reduction by allocating capacities each time period according to the shared information, compared to stable staffing in serial service operations.

The objective of our model is to minimize total cost which is the sum of labor cost and waiting cost. These two kinds of costs have been considered in many service-scheduling solutions (Mabert 1979; Grassman 1988; Davis 1991, Ittig 1994). Mohan et al. (2014) also argued that maintaining sufficient server capacity is important to guarantee a desired quality of service. Therefore, penalty (waiting) cost needs to be taken into account in the models.

The next section describes the model setup and the two capacity-allocating policies to be compared. Section 3 shows the results of numerical experiments, and Section 4 concludes the paper with managerial implications.

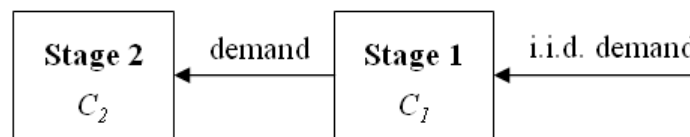
2. MODEL

Suppose there are independent and identically distributed demands for serial service operations as in Figure 1.

C_1 = the capacity of stage 1, or the number of servers at stage 1

C_2 = the capacity of stage 2, or the number of servers at stage 2

Figure 1. Model setup



If the sum of the servers in the two stages is fixed as $C_1 + C_2 = C$, how can we allocate the number of servers to each stage? If the two stages share the demand information, how can it help to allocate the capacity?

We assume that the time periods are discrete, and one period stands for service time as well as customer inter-arrival time. It is usually the case since most service organizations take certain duration of time slots for services. At the beginning of every period, a bundle of customers arrive at stage 1 and another bundle of customers who are given services from stage 1 arrive at stage 2. During each period, customers are served at each stage as the capacity allows. One server can serve one customer at one time period. Customers who cannot be served due to the capacity limit form a queue in front of each stage waiting for being served for the next period. They have the priority to be served at the next period. At the end of each period, the penalty (waiting) costs from the queues are realized and the capacity decision is made. New capacity allocation applies to the beginning of the next period.

With this sequence of events, we specify two policies – a basic policy and an information-based policy. The basic policy means that the firm allocates the half of its employees to each stage all the time. On the other hand, the information-based policy means that the firm observes the number of customers at each stage every period and decides the capacity allocation according to the information. Therefore, the major assumption under this policy is that the servers should be trained for both jobs so that they can be assigned any time needed. The problems under each policy are formulated as follows.

C = total capacity (total number of servers)

C_{it} = the capacity (number of servers) at stage i in period t

x_{it} = number of customers being served or in the queue at stage i in period t

p = penalty cost (waiting cost) per customer per period

l = labor cost per employee per period

m = the mean of demand

s = the standard deviation of demand

q_{it} = number of customers in the queue at stage i in period t

$V_t(x_{1t}, x_{2t})$ = total cost of the remaining periods from period t

Basic policy:

$$V_t(x_{1t}, x_{2t}) = \min_{C_{1t}, C_{2t} \in \mathbb{Z}^+} \{ p[(x_{1t} - C_{1t})^+ + (x_{2t} - C_{2t})^+] + lC + E_{D_{t+1}}[V_{t+1}((x_{1t} - C_{1t})^+ + D_{t+1}, \min(C_{1t}, x_{1t}))] \}$$

Information-based policy:

$$V_t(x_{1t}, x_{2t}) = \min_{C_{1t}, C_{2t} \in \mathbb{Z}^+} \{ p[(x_{1t} - C_{1t})^+ + (x_{2t} - C_{2t})^+] + lC + E_{D_{t+1}}[V_{t+1}((x_{1t} - C_{1t})^+ + D_{t+1}, \min(C_{1t}, x_{1t}))] \}$$

We first had to choose the information-based policy among some possible candidates. We designated three policies for possible candidates and found out that the following policy produces the best outcome among them with an extensive simulation study:

$$C_{2(t+1)} = \min(C_{1t}, x_{1t}) \text{ and } C_{1(t+1)} = C - C_{2(t+1)}.$$

The other two were:

$$\begin{aligned} 1) \quad C_{1(t+1)} &= \left\lceil C \times \frac{x_{1t} + m}{x_{1t} + x_{2t} + 2m} \right\rceil \text{ and } C_{2(t+1)} = C - C_{1(t+1)}; \\ 2) \quad C_{1(t+1)} &= \left\lceil C \times \frac{q_{1t} + m}{q_{1t} + q_{2t} + 2m} \right\rceil \text{ and } C_{2(t+1)} = C - C_{1(t+1)}, \text{ and } \lceil \cdot \rceil \text{ is the function that generates the nearest integer.} \end{aligned}$$

Therefore, we decide the information-based policy for this study as $C_{2(t+1)} = \min(C_{1t}, x_{1t})$ and $C_{1(t+1)} = C - C_{2(t+1)}$ that is, we assign servers to stage 2 with the exact number of customers who came out of the first stage in the previous period. The notable advantage of this policy is that it removes uncertainty in demand at stage 2. Though the waiting cost at stage 1 may increase, the total cost is lower because of no waiting in stage 2.

3. COMPUTATIONAL STUDY

In this section, we compare the total cost (customer waiting cost + labor cost) of the two policies to observe the benefit of information sharing. With the following combination of parameters, we first find the optimal capacity for each policy and then run the simulation program to calculate the costs. The number of simulated periods is 10,000 which stands for the number of time slots per day. The customer demands are assumed to be independent and identically distributed, and follow truncated Normal distributions.

$$m = \{10, 20, 30\};$$

$$S = \{0.1m, 0.2m, 0.3m\};$$

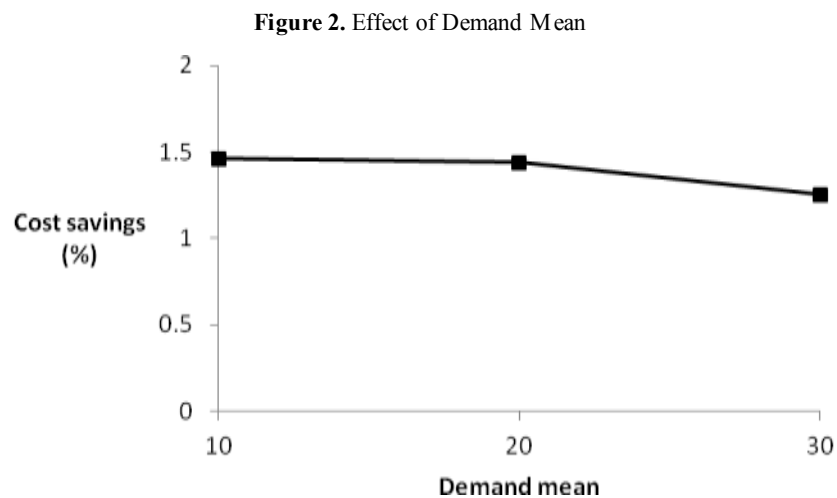
$$l = 1 \text{ for all experiments (fixed as a ground measure); and}$$

$$p = \{0.5, 1, 1.5, 2\}.$$

We detect the cost difference by comparing the two policies under the computational setting. The percentage difference is calculated as follows.

$$\text{Cost savings (\%)} = 100 \cdot \frac{\text{average cost under basic policy} - \text{average cost under information based policy}}{\text{average cost under basic policy}}$$

This measure indicates the cost savings by moving from the basic policy to the information-based policy. As a result of simulation, we can say the information-based policy performs always better than the basic policy ranged from 0.296% to 3.613% with an average of 1.385%. The cost difference seems relatively not so large. Considering that the information-based policy requires significant cross-training for servers, the basic policy is quite an effective policy as it is.



From the Figure 2, the effect of demand mean seems negligible considering that demand distributions are truncated Normal.

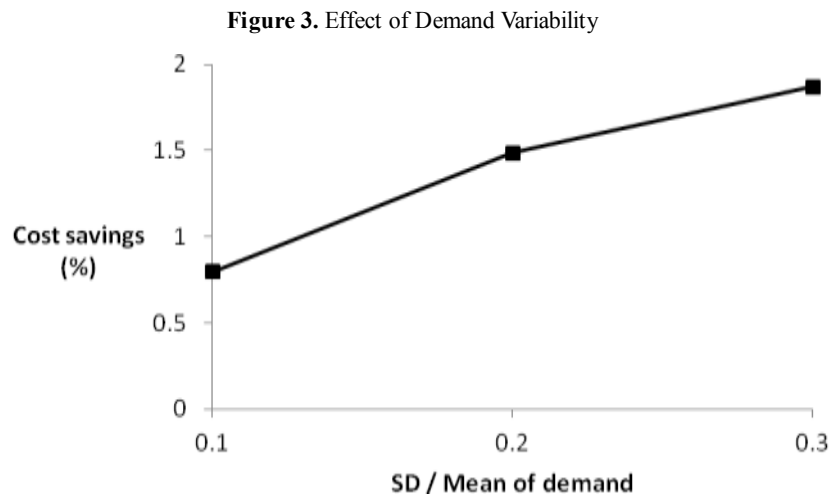


Figure 3 shows the impact of demand variability. We can observe that the value of information sharing increases as the ratio of standard deviation to mean of the demand. As the demand standard deviation increases, the degree of demand uncertainty at each stage is also increasing and the benefit of uncertainty reduction at stage 2 seems increasing as well.

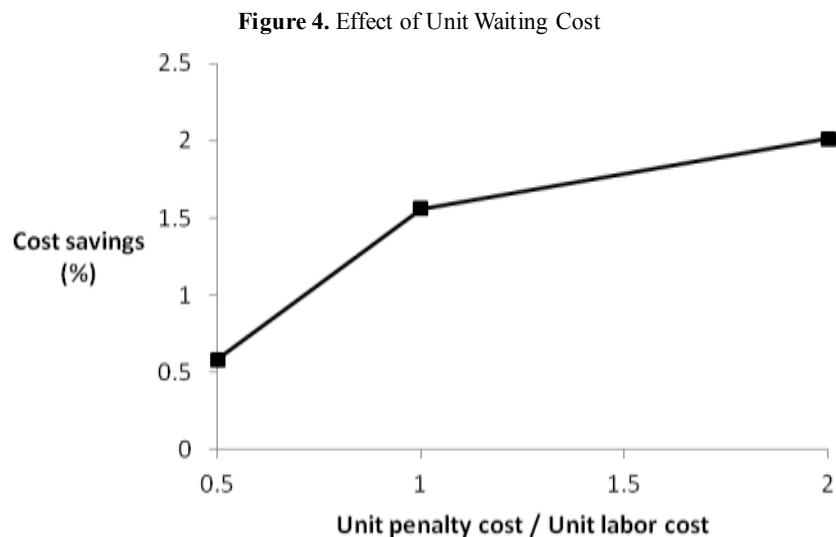


Figure 4 illustrates the impact of unit penalty (waiting) cost on cost savings when the unit labor cost is fixed as a ground measure. We observe that the value of information sharing increases with the unit penalty cost. If waiting is more expensive, customers at stage 2's not waiting can be more valuable.

4. CONCLUSION

The value of information sharing in manufacturing supply chains has been well proved and the value comes from reducing demand uncertainties faced by suppliers. Similarly, for serial service operations, if we utilize the demand information to allocate the labor capacities in two stages, we can reduce the total service cost including waiting and labors.

As an information-based policy, we used a rule to assign the same number of servers as that of the customers who came out of the first stage in the previous period. This policy is observed to perform the best among several possible policies of information sharing, and has a benefit of removing demand uncertainties at the latter stage of the two.

By conducting a simulation study with various parameters, we found out that there exists the value of information sharing in capacity allocation in serial service operations though the percentage cost savings are not so high. We have also observed that the value of information increases with demand variability or with unit waiting cost.

Considering that the information-based policy requires significant amount of cross-training for employees, we have to weigh the benefit of the policy and the training cost. As the cost savings from using the information-based policy are relatively low, the basic policy may be quite an effective policy. The basic policy is simple to implement and does not require further training of employees.

From the computational study, we observe that the cost savings out of using the information-based policy over the basic policy increase as the demand variability increases or as the unit penalty (waiting) cost increases. The results imply that information sharing in serial service operations can be more valuable in such environment as demand is highly variable or as waiting is more expensive.

For future research, we may extend the analysis to more combinations of parameters in the simulation study. For example, we can include more demand distributions such as Uniform and Exponential or more cost parameters. It would be also interesting to observe whether the optimal numbers of total servers are different by policies. For the information-based policy, we adapt an assigning rule such that the number of servers in the next stage is determined by the number of customers who have been served at the earlier stage in the previous period. However, it has not been mathematically proved to be optimal. If there exists an optimal policy for information sharing in the serial service operations, it would be a very important subject to search for such a policy. Finally, we may extend our analysis to the case of three-stage service operations. As this is the first study to investigate the value of information sharing in serial service operations, it is expected that there can be lots of possible future research regarding this subject.

AUTHOR BIOGRAPHY

Jin Kyung Kwak graduated from Seoul National University in South Korea, majoring in Business Administration and obtained the bachelor's and master's degree. She got her Ph.D. from the Johnson Graduate School of Business, Cornell University in Ithaca, New York. Her research interests lie in the supply chain management area. Since 2012, She has been an assistant professor in Ewha School of Business in South Korea.

REFERENCES

- Alfares, H.K. (2004), "Survey, Categorization, and Comparison of Recent Tour Scheduling Literature," *Annals of Operations Research*, 127, 145-175.
- Browne, J. (1997), "A Framework for Service Employee Scheduling," *Industrial Management*, May/June, 1-2.
- Cappelli, P. (2009), "A Supply Chain Approach to Workforce Planning," *Organizational Dynamics*, 38, 1, 8-15.
- Davis, M.M. (1991), "How Long Should a Customer Wait for Service," *Decision Sciences*, 22, 2, 421-434.
- Gavirneni, S., R. Kapuscinski, S. Tayur (1999), "Value of information in capacitated supply chains," *Management Science*, 45, 1, 16-24.
- Grassman, W.K. (1988), "Finding the Right Number of Servers in Real-World Queuing Systems," *Interfaces*, 18, 2, 94-104.
- Ittig, P.T. (1994), "Planning Service Capacity When Demand is Sensitive to Delay," *Decision Sciences*, 25, 4, 541-559.
- Lee, H.L., K.C. So, C.S. Tang (2000), "The value of information sharing in a two-level supply chain," *Management Science*, 46, 5, 626-643.
- Mabert, V.A. (1979), "A Case Study of Encoder Shift Scheduling Under Uncertainty," *Management Science*, 25, 7, 623-631.
- Mohan, S., F.M. Alam, J.W. Fowler, M. Gopalakrishnan, A. Printezis (2014), "Capacity Planning and Allocation for Web-Based Applications," *Decision Sciences*, 45, 3, 535-567.
- Song, H., H-C Huang (2008), "A successive convex approximation method for multistage workforce capacity planning problem with turnover," *European Journal of Operational Research*, 188, 29-48.
- Zhu, X., H.D. Sherali (2009), "Two-stage workforce planning under demand fluctuations and uncertainty," *Journal of the Operational Research Society*, 60, 94-103.